<u>PATENT APPLICATION ENTITLED</u>:

SYSTEM AND METHOD FOR DYNAMIC BANDWIDTH PROVISIONING

<u>BY INVENTORS</u>:

TUAN TA

PATRICK TURLEY

KERRY CLENDINNING

KELLY LOONEY

<u>DATE FILED</u>:

OCTOBER 16, 2003

<u>ATTORNEY DOCKET NO.</u>:

ROCK1120-1

<u>ATTORNEYS FOR APPLICANTS</u>

Customer No. 25094
Gray Cary Ware & Freidenrich LLP
1221 South MoPac Expressway, Suite 400
Austin, TX 78746-6875
Attn: John L. Adair
Tel. (512) 457-7142
Fax. (512) 457-7001

<u>EXPRESS MAIL LABEL NO.</u>:

EV351127145US

System and Method For Dynamic Bandwidth Provisioning

Related Applications:

[0001] This application claims priority under 35 U.S.C. 119(e) to United States Provisional Patent Application No. 60/418,968, entitled "System and Method for Dynamic Bandwidth Shaping," to Kerry Clendenning, et al., filed October 16, 2002, which is hereby fully incorporated by reference herein.

TECHNICAL FIELD OF THE INVENTION

[0002] Embodiments of the present invention relate to providing network access. More particularly, embodiments of the present invention relate to provisioning bandwidth to users accessing a network. Even more particularly, embodiments of the present invention relate to dynamically provisioning bandwidth on a per user basis.

BACKGROUND

[0003] The communication of data over networks has become an important, if not essential, way for many organizations and individuals to communicate. The Internet is a global network connecting millions of computers using a client-server architecture in which any computer connected to the Internet can potentially receive data from and send data to any other computer connected to the Internet. The Internet provides a variety methods in which to communicate data, one of the most ubiquitous of which is the World Wide Web. Other methods for communicating data over the Internet include e-mail, usenet newsgroups, telnet and FTP.

[0004] Users typically access the Internet either through a computer connected to an Internet Service Provider ("ISP") or computer connected to a local area network ("LAN") provided by an organization, which is in turn, connected to the ISP. The ISP provides a point of presence to interface with the Internet backbone. Routers and switches in the backbone direct data traffic between the various ISPs.

[0005] To access a LAN and, in turn, the Internet, many prior art access control systems require a user to connect his or her computer to a wired network (e.g., through an Ethernet port) and enter a user name and password. If the user name and password match a user name and password in an authentication database, the user will be provided

access to the network. These systems typically assume that a user is tied to a particular physical port, such as a port in the user's office. Based on this assumption, provisioning of bandwidth to the user occurs by physically provisioning the port to which the user is connected. If the user moves to a different port, the user will typically be provided with the bandwidth provisioned to the new port. Thus, provisioning of bandwidth is done on a per port rather than a per user basis.

[0006] An increasing number of organizations (e.g., businesses, governmental organizations) wish to provide access to LANs and the Internet to various classes of users (internal users, contractors, customers, visitors). For example, many cafés have public wireless networks to allow patrons to access the Internet, receive email and perform other network activities. While users may be asked to authenticate to use the network, bandwidth is provisioned to the wireless routers, not the individual users. This means that one user connected to a particular router can consume a majority of the bandwidth (e.g., downloading pictures from the Internet), slowing down the wireless network for other users connected to that router.

[0007] Because bandwidth is provisioned on a per port rather than per user basis, current systems can not reprovision bandwidth to particular users as more users are added to the same port (e.g., as more users connect to the same

wireless router). Continuing with the previous example of a public wireless LAN at a café, current systems can not reprovision bandwidth to individual users as more café patrons connect to the LAN. This can cause LAN performance to suffer for all users.

SUMMARY OF THE INVENTION

[0008] Embodiments of the present invention provide a system and method of providing bandwidth allocation that eliminates, or at least substantially reduces, the shortcomings of prior art bandwidth provisioning systems and methods. More particularly, one embodiment of the present invention provides a device for allocating bandwidth on a per user basis. The device can be a computing device that comprises a processor, a first network interface (e.g., an Ethernet interface, T1 interface, wireless interface or any other network interface known in the art) coupled to the processor, a second network interface (e.g., an Ethernet interface, T1 interface, wireless interface or any other network interface known in the art) coupled to the processor, and a storage medium (e.g., hard disk drive, RAM, ROM, optical storage and or any other storage medium known in the art). The storage medium can be accessible by the processor and can contain a set of computer instructions.

[0009] The computer instructions, according to one embodiment of the present invention, can be executable by the processor to retrieve a set of user profiles, wherein each user profile corresponds to a specific user in a set of users. Each user profile can contain an arbitrary number of attributes that can specify bandwidth limitations for the corresponding user. Based on the user profile for each user, the computer instructions can be executable to establish at least one bandwidth limit for each user and

control bandwidth usage by that user accordingly. The computer instructions can also be executable to update the at least one bandwidth limit for one or more users.

[0010] Another embodiment of the present invention can include computer instructions stored on a computer readable medium (e.g., RAM, ROM, hard disk drive, magnetic storage device, optical storage device or other computer readable medium known in the art). The computer instructions can be executable by a processor to retrieve a set of user profiles, wherein each user profile corresponds to a specific user in a set of users. Each user profile can contain an arbitrary number of attributes that can specify bandwidth limitations for the corresponding user. Based on the user profile for each user, the computer instructions can be executable to establish at least one bandwidth limit for each user. For each user in the set of users, the computer instructions can be further executable to regulate bandwidth usage associated with that user based on the at least one bandwidth limit established for that user. The computer instructions can also be executable to update the at least one bandwidth limit for one or more of the users.

[0011] Yet another embodiment of the present invention can include a method comprising retrieving a set of user profiles, wherein each user profile corresponds to a specific user in a set of users; establishing at least one bandwidth limit for each user in the set of users based on the corresponding user profile for that user;

for each user in the set of users, regulating bandwidth usage associated with that user based on the at least one bandwidth limit established for that user; and updating the at least one bandwidth limit for at least one user from the set of users.

[0012] Yet another embodiment of the present invention comprises computer instructions stored on a computer readable medium (e.g., RAM, ROM, hard disk drive, magnetic storage device, optical storage device or other computer readable medium known in the art). The computer instructions can be executable by a processor to establish a bandwidth limit for a user based on a user profile for the user; receive a first network communication; determine if the first network communication causes the bandwidth limit to be exceeded; if the first network communication causes the bandwidth limit to be exceeded, drop the network communication; and update the bandwidth limit for the user.

[0013] Embodiments of the present invention provide an advantage over prior art bandwidth allocation systems and methods by allowing bandwidth to be provisioned and dynamically updated on per user basis.

[0014] Embodiments of the present invention provide another advantage over prior art bandwidth allocation systems and methods by allowing bandwidth usage to be metered on a per user basis.

[0015] Embodiments of the present invention provide another advantage by allowing users to be allocated a predefined amount of bandwidth on wired and wireless networks.

BRIEF DESCRIPTION OF THE FIGURES

[0016] A more complete understanding of the present invention and the advantages thereof may be acquired by referring to the following description, taken in conjunction with the accompanying drawings in which like reference numbers indicate like features and wherein:

[0017] FIGURE 1 is a diagrammatic representation of a system for providing bandwidth allocation according to one embodiment of the present invention;

[0018] FIGURE 2 is a flow chart illustrating one embodiment of a method for providing per user bandwidth allocation;

[0019] FIGURE 3 is a diagrammatic representation of one embodiment of a software system for providing per user bandwidth allocation;

[0020] FIGURE 4 is a diagrammatic representation of traffic conditioning module, according to one embodiment of the present invention;

[0021] FIGURE 5 is a diagrammatic representation of a system for dynamically allocating bandwidth to users on a per user basis, according to one embodiment of the present invention;

[0022] FIGURES 6A-6D illustrate various embodiments of bandwidth shaping; and

[0023] FIGURE 7 is a diagrammatic representation of one embodiment of a control device that can provide user based provisioning of bandwidth.

DETAILED DESCRIPTION

[0024] Preferred embodiments of the invention are illustrated in the FIGURES, like numerals being used to refer to like and corresponding parts of the various drawings.

[0025] Embodiments of the present invention provide a system and method for providing bandwidth shaping on a per user basis. For purposes of this application, "bandwidth shaping" refers to the process of determining bandwidth allocations. According to one embodiment of the present invention, a control device can be located between a network (e.g., the Internet, a LAN or other network known in the art) and users. The users themselves may be located on a network or subnet or may connect directly to the control device. When a particular user attempts to access the network behind the control device (i.e., the controlled network), the control device can assign a bandwidth to the user based on a user profile. As more users attempt to access the controlled network through the control device, the control device can dynamically alter the bandwidth allocations among users. Additionally, the control device can reallocate bandwidth on a per user basis to account for various factors, such as time of day, burstiness or other such factors known in the art.

[0026] By way of example, but not limitation, a control device can be coupled to a public wired or wireless LAN and the Internet to provide users with access to the Internet.

When a user attempts to access the Internet over the public LAN, the control device can retrieve a user profile for the user and assign a bandwidth to the user. As additional users attempt to access the Internet, the control device can reassign bandwidths to the user.

[0027] Embodiments of the present invention provide an advantage over prior art bandwidth shaping systems and methods by allocating bandwidth on a per user basis. Additionally, embodiments of the present invention provide an advantage by allowing the bandwidth allocations to be dynamically altered to account for, for example, changes in network state, user activity or number of users. Embodiments of the present invention can allocate and dynamically shape bandwidth for both wired and/or wireless networks.

[0028] FIGURE 1 is a diagrammatic representation of a system 10 for providing bandwidth allocation according to one embodiment of the present invention. In system 10, a user 12, using a user device 14 on network 16, can send a network communication that is destined for a device on a controlled network 18. It should be noted that a user can be a human user, a programmatic user, or other user. User device 14 can comprise a desktop, a laptop, a PDA, a cell phone, a desktop computer or any other computing device capable of network communications. Network 16 and network 18 can be any networks known in the art including, but not limited to, LANs, WANs, the Internet, global communications networks, wireless networks and/or any other communications networks known in the art. For

the sake of example, network 16 can be a wireless network and controlled network 18 can be the Internet. An access control device 20 ("control device 20") can control access to network 18 by users of network 16.

[0029] According to one embodiment of the present invention, when user 12 connects to control device 20 (e.g., by sending a network communication destined for network 18 from user device 14), control device 20 can retrieve a user profile for user 12 that governs the bandwidth allocations to user 12. In one embodiment of the present invention, control device 20 can receive the user profile for user 12 from an authentication database 24 at an authentication system 22 based on a set of user credentials provided by user 12 (e.g., password, user id, biometric data, certificates and or other user provided or hardware credentials known in the art). In another embodiment of the present invention, the user profile for user 12 can be stored at control device 10 in, for example, a local authentication database. Example authentication technologies include LDAP, Remote Authentication Dialup Internet User Specification ("RADIUS"), Microsoft Active Directory Service, Tivoli Access Manager, and Cisco TACACS/TACACS+. Authentication can occur as described in United States Patent Application Serial No.__________, entitled "System and Method for Providing Access Control," filed October 10, 2003, to MacKinnon et al., which is hereby fully

incorporated by reference herein (the "Access Control Application").

[0030] Based on the user profile, control device 20 can establish a traffic control rule specifying maximum upload and download bandwidths for user 12. When network communications are received from user 12, control device 20 can determine if the maximum upload bandwidth has been exceeded and, if so, drop or queue the communication. When communications to user 12 are received from controlled network 18, control device 20 can determine if the maximum download bandwidth has been exceeded and, if so, drop or queue the communication. If the bandwidth allocations are not exceeded, control device 20 can pass the communication to/from user 12.

[0031] FIGURE 2 is a flow chart illustrating one embodiment of a method for providing per user bandwidth allocation. In one embodiment of the present invention, a control device (e.g., control device 20 of FIGURE 1) can authenticate a user (step 26) based on a network communication from a network application such as a web browser, ftp client, telnet client, mail user agent or other publicly available or proprietary network application running on a user device (e.g., laptop 14). Authentication can occur in any manner known in the art and, in one embodiment, can be carried out by the control device as described in the Access Control Application.

[0032] At step 28, the control device can retrieve a user profile. The user profile can be retrieved from remote storage (e.g., a remote authentication system) or from local storage (e.g., a local authentication database). The user profile can contain attributes that govern bandwidth allocation to an associated user. For example, the user profile can contain indicators of the upload and download bandwidth to which the user is entitled. The user profile can be received in a canonical format as an HTTP Post from the authentication system to the control device or in any manner known in the art.

[0033] According to one embodiment of the present invention, each user profile can have a predefined set of attributes. In some cases, the retrieved profile may not provide values for each of these attributes. Therefore, the control device can determine if the received user profile is complete, and, if it is not complete, can fill in the missing attribute values with default values, which can be part of the control device's local configuration or may be retrieved from remote storage by the control device during, for example, its initialization or startup phase.

[0034] At step 30, the control device can initiate a control session for the authenticated user. The control session can have a specified time limit, a time out limit and/or other session features known in the art. While the control session is active, the control device can govern

a user's access to a network according to the user profile associated with that user and defaults and monitor the control session for characteristics such as bandwidth used or network session characteristics known in the art. The control session can be tracked by credentials, IP address, MAC address and/or other identifier.

[0035] At step 32, the control device can establish user specific rules and conditions based on attributes in the user profile. According to one embodiment of the present invention, the control device can map the attributes to a traffic control rule. By way of example, but not limitation, a user profile can contain attributes to specify upload and download bandwidth allocations for a user that can be mapped to a traffic control rule usable by a traffic conditioning module, such as the Linux based Traffic Control application, to regulate bandwidth usage by the corresponding user.

[0036] In one embodiment of the present invention, rules can be represented in an IP table. As would be understood by those of ordinary skill in the art, IP tables are essentially tables of rules that can be accessed by applications, such as a firewall or other application configured to access the IP tables. Rules in the IP table can be associated to a user through any arbitrary identifier. Using the example of FIGURE 1, the IP table rule(s) for user 12 can be bound to user 12 based on the MAC address and IP address of user device 14 and the

credentials provided by user 12 for the particular control session. If, for example, laptop 14 is associated with the MAC address 08:00:69:02:01:FF and the IP address 100.100.100, the user specific rules for user 12 can be indexed under these addresses and, for example, a user name for the control session.

[0037] An IP table rule can reference other rules or parameters for providing user specific provisioning. As an example, the IP table rule for user 12 can reference a traffic control rule that dictates that the bandwidth allocated to user 12 is 128 Kbps for uploads and 512 Kbps for downloads. A traffic control application, such as the Linux based Traffic Control application, can access the traffic control rule through the IP table and enforce the bandwidth allocation.

[0038] It should be noted that a user specific rule in an IP table can reference rules or parameters usable by any number of applications or processes, such as firewalls, traffic conditioning modules, virus scan applications or other applications known in the art. For example, a user specific rule can contain or reference parameters usable by a virus scan application to provide user specific virus scanning. It should be further noted that the use of IP tables to index user specific rules for provisioning of bandwidth is provided by way of example only, and user specific rules can be implemented in any suitable manner, as would be understood by those of ordinary skill in the art.

[0039] The control device, at step 34, can receive a network communication to/from a particular user. The network communication can include, by way of example, but not limitation, an HTTP message, an email message, a telnet message, an FTP message, UDP message or other network communication known in the art. As would be understood by those of ordinary skill in the art, many network communications (e.g., HTTP requests/responses, email messages, instant chat messages and other network communication known in the art) are carried by one or more IP packets. Each IP packet includes a header with originating IP address, destination IP address, originating MAC address, destination MAC address and/or other addressing information.

[0040] At step 36, the control device can determine which user specific rule(s) to apply to the network communication. This can be done, for example, by extracting information from IP packets associated with the network communication. For a network communication from a particular user, the control device can compare the originating IP address and/or MAC address to the IP addresses and/or MAC addresses used to index user specific rules in the IP table. For a network communication to a user, the control device can compare the destination IP address and/or MAC address to the IP addresses and/or MAC addresses used to index user specific rules in the IP table. Based on this

information, the control device can access the appropriate user specific traffic control rule(s).

[0041] In one embodiment of the present invention, a traffic control rule can specify user specific upload and download bandwidths. The control device can apply the traffic control rule(s) for a user to regulate bandwidth usage by the user. Thus, the control device, at step 38 can determine if a network communication causes the maximum upload bandwidth limit for an originating user or the maximum download bandwidth for a destination user to be exceeded. If appropriate bandwidth limit is exceeded, the control device can queue or drop the network communication (step 40). If, on the other hand, the bandwidth limit is not exceeded, the control device can transmit the network communication to its destination (step 42). This process can be repeated (step 44) for each network communication.

[0042] FIGURE 3 is a diagrammatic representation of one embodiment of a software system for providing per user bandwidth allocation. According to one embodiment of the present invention, a set of computer instructions 48 running at control device 20 can include a provisioning module 50 to provision or allocate bandwidth to users. Provisioning module 50, according to one embodiment of the present invention can establish rules associated with a particular user based on a user's profile. For example, provisioning module 50 can establish user specific rule 52 for user 12 based on user profile 54.

The user specific rules, in one embodiment of the present invention can be indexed through an IP table 56. Each user specific rule can be associated with a user based on an arbitrary identifier. For example, IP table 56 can contain user specific rule 52 for user 12 indexed to MAC address 08:00:69:02:01:FF and the IP address 100.100.100 of user device 14.

[0043] Each user specific rule in IP table 56 can optionally point to additional rules and parameters. For example user specific rule 52 can point to traffic control rule 58, which can also be based on user profile 54, to govern bandwidth provisioning to user 12. For example, user profile 54, in one embodiment of the present invention, can specify that user 12 is entitled to 128 Kbps for uploads and 512 Kbps for downloads. Provisioning module 50 map these attributes to a traffic control rule that contains these limitations and is usable by traffic conditioning module 60 to regulate bandwidth usage. One example of a traffic conditioning module that can be configured to access traffic control rules through an IP table is the Linux based Traffic Control application, however, any known or proprietary traffic conditioning set of computer instructions can be used by various embodiments of the present invention to apply per user traffic control rules.

[0044] In operation, provisioning module 50 can receive an IP packet 62 from user device 14. In one embodiment of the present invention, traffic conditioning module 60 can

extract the originating IP address and/or MAC address from IP packet 62, access IP table 56 and locate user specific rule 54 on the originating MAC address and IP address. From user specific rule 52, traffic conditioning module 60 can locate traffic control rule 58 that specifies a maximum upload bandwidth of 128 Kbps. If packet 62 causes this limit to be exceeded, traffic conditioning module 60 can queue or drop packet 62. Otherwise, traffic conditioning module 60 can communicate packet 62 to its destination. Packet flow in the reverse direction can be processed in an analogous manner. For example, a packet arriving from the Internet destined for a user device 14 can be examined for destination IP address and MAC address. Based on these identifiers, traffic conditioning module 60 can access and apply traffic control rule 58 to limit the download bandwidth to 512 Kbps.

[0045] In the above example, user profile 54 contained attributes to govern the upload and download bandwidth allocations to user 12 regardless of the network application generating or receiving the IP packets. In another embodiment of the present invention, user profile 54 can prioritize network applications. As an example, user profile 54 can specify that IP packets associated with a game have priority over email messages, which have priority of HTTP messages. The priorities can be mapped to user specific traffic control rule 58. Continuing with the previous example, for the network application

that is assigned a highest priority, traffic control rule 58 can specify that that network application is entitled to 320 Kbps for downloads and 80 Kbps for uploads, the network application assigned the next highest priority is entitled to 120 Kbps for downloads and 32 Kbps for uploads, and the network application assigned the lowest priority is entitled to maximum download and upload bandwidths of 64 Kbps and 16 Kbps, respectively. Traffic conditioning module 60 can read each IP packet to determine the network application with which it is associated, using any application detection scheme known in the art, and apply the application specific bandwidth limitations for user 12 in traffic control rule 58.

[0046] In another embodiment of the present invention, if user 12 exceeds the user specific bandwidth limits (e.g., 128 Kbps for uploads or 512 Kbps for downloads), traffic conditioning module 60 can simply drop packets associated with lower priority network applications until the upload or download bandwidth usage falls within the specified limit rather than assigning specific bandwidth limits to each application or class of applications. The present invention can also employ any other prioritization scheme as would be understood by those of ordinary skill in the art.

[0047] In addition to provisioning bandwidth to users on a per user basis, the control device can provide session monitoring at, for example, monitoring module 61. Monitoring module 61 can perform any session monitoring

functions known in the art, including, but not limited, tracking session time, tracking timeouts, generating reports and metering bandwidth usage on a per session basis. One example of a monitoring module can be a Radius client. It should be noted that because bandwidth allocation occurs on a per user basis bandwidth metering can also occur on a per user basis. Any bandwidth metering scheme known in the art can be used.

[0048] FIGURE 4 is a diagrammatic representation of traffic conditioning module 60, according to one embodiment of the present invention. Conditioning module 60 can include interface master queue 68, user discriminator 70 and user specific conditioners 72. User discriminator 70 can read a packet header to determine the appropriate user specific conditioner that will process the packet based on, for example, the IP address and MAC address of the packet. The user conditioner 72 can access the appropriate traffic control rule from IP table 56 to enforce user specific traffic conditions (i.e., upload and download bandwidth limits).

[0049] According to one embodiment of the present invention, conditioning module 60 can locate the user specific traffic control rule based on user specific IP table 56. For example, for a packet having the originating MAC address 08:00:69:02:01:FF and the IP address 100.100.100, traffic conditioning module 60 can access user specific rule 52 and, from user specific rule 52, user specific traffic control rule 58. The user specific traffic

control rule can be enforced at the corresponding user specific conditioner.

[0050] For a particular network interface, each user can have an inward (i.e., download) and outward (i.e., upload) bandwidth allowance, based on attributes in the user's user profile. A bandwidth limit is the maximum rate at which a user is permitted to transmit or receive network traffic over a particular interface. User specific traffic conditioners 72 ensure that, if a user exceeds his or her bandwidth allowance for that interface, further network traffic for the user in that direction will be queued or dropped until the average data rate falls back within the bandwidth allowance. Thus, traffic conditioning module 60 can regulate bandwidth on a per user basis.

[0051] Interface master queue 68 can control the flow of network traffic over a particular interface (e.g., a particular Ethernet interface, wireless interface T1 interface, or other interface known in the art). It can be configured to send out data at whatever rate is appropriate for the corresponding network connection. Interface master queue 68 can be feed IP packets by user specific conditioners 72, each of which can have its own queue to hold packets the conditioner has accepted, but interface master queue 68 is not ready to accept.

[0052] Interface master queue 68, according to one embodiment o the present invention, can also reallocate bandwidth to

particular users based on excess capacity, usage patterns, time of day, user priorities or other factors known in the art. As an example, when bandwidth is available due to low demand by other users, a particular user may achieve rates in excess of the limits specified in the user's user profile. Excess bandwidth can be divided "fairly" between subscribers. When demand exceeds the available bandwidth, the bandwidth allocated to a user can throttled to or below the amount of bandwidth designated by the user profile.

[0053] It should be noted that the architectures of FIGURE 3, and FIGURE 4 are provided by way of example only and bandwidth provisioning and monitoring can be implemented using any suitable programming language and/or structure known in the art. In the example of FIGURE 3, provisioning module 50 provides traffic conditioning based on a set of IP tables. It should be noted, however, that the rules for traffic conditioning can be defined in any arbitrary manner usable by a traffic conditioning module. It should be further noted, that provisioning module 50 can provide additional services on a per user basis, such as virus scanning, worm detection, firewall services or any other service known in the art. Each of these additional services can access rules and parameters based, for example, on IP table(s). Moreover, the use of IP tables to index and provide access to rules and parameters for particular processes is also provided by way of example. In other embodiments of the present

invention, a user profile can be mapped to rules and parameters for various applications in any suitable programming manner known in the art.

[0054] In the example of FIGURE 3, traffic conditioning module 60 applied the bandwidth limitations in traffic control rule 58 as hard limits. That is, regardless of overall available bandwidth, traffic conditioning module 60 will limit the upload and download bandwidths to 128 Kbps and 512 Kbps limits specified in traffic control rule 58. In another embodiment of the present invention, the traffic conditioning module at, for example, interface master queue 68, can alter the bandwidth limits for a user dynamically to account for excess capacity, usage patterns, number of user, or any other factor.

[0055] FIGURE 5 is a diagrammatic representation of a system for dynamically allocating bandwidth to users on a per user basis, according to one embodiment of the present invention. In FIGURE 5, user 80 (User A) using user device 82, user 84 (User B) using user device 86 and user 88 (User C) using user device 90 on network 92 can send a network communications destined for locations on a controlled network 94. Network 92 and network 94 can be any networks known in the art including, but not limited to, LANs, WANs, the Internet, global communications networks, wireless networks and/or any other communications networks known in the art. For the sake of example, network 92 can be a wireless network and controlled network 94 can be the Internet. A control

device 96 can control access by users on network 92 to controlled network 94 and allocate bandwidths to the users on a per user basis.

[0056] In one embodiment of the present invention, control device 96 can retrieve a user profile for each user on network 92 (e.g., user profile 98 for User A, user profile 100 for user B and user profile 102 for User C). For each user profile, a provisioning module 104 at control device 96 can establish a user specific rule (user specific rule 106 for User A, user specific rule 108 for User B and user specific rule 110 for User C) indexed, for example, in IP table 112. Additionally, based on the user profiles, provisioning module 104 can establish traffic control rules for each user (e.g., traffic control rule 114 for User A, traffic control rule 116 for User B and traffic control rule 118 for User C) which can be referenced by the respective user specific rule in IP table 112. Each traffic control rule can be based on the user profile and can specify bandwidth allocations for the corresponding user. It should be noted that the use of one or more IP tables to index traffic control rules through user specific rules is provided by way of example only, and the present invention can specify bandwidth allocations on a per user basis in any suitable manner.

[0057] In operation, traffic conditioning module 120 can receive packets to/from User A, User B and/or User C and process the packets according to the corresponding traffic

control rules. For example, traffic conditioning module 118 can receive a packet from User A, access user specific rule 106 in IP table 112 based, for example, on a MAC address and IP address for user device 82, access traffic control rule 114 referenced by user specific rule 106, and apply traffic control rule 114 to drop the packet, queue the packet or communicate the packet to network 94 based on whether or not the packet causes bandwidth limits in traffic control rule 114 to be exceeded.

[0058] Each user profile, in one embodiment of the present invention, can specify that a corresponding user is entitled to a particular bandwidth. Provisioning module 104 can establish traffic control rules that contain these limitations. However, the traffic control rules if simply based on the user profiles may not account for a variety of network conditions, such as, for example excess capacity, new users, usage conditions and other such factors.

[0059] As an example, assume control device 96 has 1000 Kbps available as its overall download bandwidth capacity, user A is allocated a maximum of 500 kbps download bandwidth based on user profile 98, user B is allocated a maximum of 250 kbps download bandwidth based on user profile 100 and only User A and User B are using network 92 to access controlled network 94 (i.e., User C is not connected to network 92). In this case, based on the user profiles, traffic control rule 114 and traffic

control rule 116 will specify maximum download bandwidths of 500 Kbps and 250 kbps respectively. If traffic conditioning module 120 applies traffic control rule 114 and traffic control 116 with download bandwidth limits of 500 kbps and 250 kbps, control device 96 will have an excess capacity of 250 kbps for downloads.

[0060] In one embodiment of the present invention, traffic conditioning module 120, at for example the interface master queue corresponding to network 94, can reassign the bandwidth limits to User A and User B to account for the excess bandwidth. For example, traffic conditioning module 120 can evenly divide the excess capacity to provide User A and User B with an additional 125 Kbps each for downloads. According to one embodiment, traffic conditioning module 120 can establish the new download bandwidth limits of 625 Kbps and 375 Kbps for User A and User B by modifying traffic control rule 114 and traffic control rule 116, respectively. Because, in this embodiment of the present invention, the traffic control rule is referenced for each user by the IP table, but is not part of the IP table, the new bandwidth limits for User A and User B can be implemented without modifying IP table 112.

[0061] As a further example, assume that based on their user profiles, User A is allocated 500 Kbps for downloads and User B has been allocated 250 Kbps for downloads, leaving an excess capacity of 250 Kbps. Assume that User C connects to network 92 to gain access to controlled

network 94. Control device 96 can retrieve user profile 102 and, based on user profile 102, establish user specific rule 110 and traffic control rule 118 for User C. If user profile 102 specifies that User C is entitled to 500 Kbps for downloads, the overall provisioned bandwidth will exceed the available bandwidth at control device 96 (i.e., the overall provisioned bandwidth will be 1250 Kbps in comparison to the available 1000 Kbps). Therefore, traffic conditioning module 120 can reallocate bandwidth among the users by modifying the traffic control rules for each user.

[0062] In one embodiment of the present invention, where there is insufficient capacity to serve all users fully, the users must essentially "compete" for what is available. If the total available bandwidth is less than the sum of all nominal allowances (e.g., the total available bandwidth of 1000 Kbps is less than the allocated bandwidth of 1250 Kbps), then the effective bandwidth offered to each user can governed by the following:

$$W = \sum w_i \qquad \text{[EQ. 1]}$$

$$e_i = a_i w_i / W \qquad \text{[EQ. 2]}$$

[0063] where: w_i is the weight applied to bandwidth allowance i for a user using an interface (e.g., the interface to network 94); W is the sum of all weights on all bandwidth allowances on this interface; e_i is the effective bandwidth for allowance i; and a_i is the nominal

bandwidth for allowance i. In the above example, User A's nominal bandwidth (a_i) allowance for using controlled network 94 is 500 Kbps. The weight w_i for each user for a given interface can be calculated based on, for example, a service factor specified in each user's user profile. If all user profiles specify the same service factor, then all users will have their bandwidth limits reduced by the same amount to bring the allocated bandwidth within the bandwidth capacity of the control device. If a user profile specifies a smaller service factor, then that user will lose a correspondingly less amount of bandwidth. If a user profile specifies that a user's bandwidth is incompressible, this means that, even if there is insufficient available bandwidth to serve all other users fully, this user's bandwidth allowance will not be reduced. This will result in more severe effective reductions for other users whose bandwidth allowances are not marked as incompressible.

[0064] In the above example, if each user profiles for User A, User B and User C specify the same service factor, traffic conditioning module 120 can reduce the maximum download bandwidth limit for each of User A, User B and User C by 83.3 Kbps (i.e., the total excess allocation of 250 Kbps divided evenly among the users). To effect this change, traffic conditioning module 120 can modify traffic control rule 114 to include a download bandwidth limit of 416.7 Kbps, traffic control rule 116 to include a download bandwidth limit of 166.6 Kbps, and traffic

control rule 118 to include a download bandwidth limit of 416.7 Kbps, for a total bandwidth allocation of 1000 Kbps among User A, User B and User C. Thus, a control device can dynamically change the allocations of bandwidth on a per user basis to account for new users.

[0065] In the above examples, traffic conditioning module 120 dynamically calculated bandwidth allocations for Users A, B and C. In another embodiment of the present invention, control device 96 can receive a new user profile for a particular user and update the traffic control rule for that user to reflect changes in the user profile. As an example, if control device 96 controls access to network 94 based on paid subscriptions by users, the initial user profile 98 for User A can specify upload and download bandwidth limits, say 250 Kbps for uploads and 500 Kbps for downloads. If User A then pays for more bandwidth during a session, user profile 98 can be updated by control device 96 or by another system (e.g., a backend system). The new user profile can now specify, for example, that User A is now entitled to 750 Kbps for downloads. Control device 96 can update traffic control rule 114 to reflect the change in maximum bandwidth for User A without requiring User A to reauthenticate to get the new bandwidth limits.

[0066] In yet another embodiment of the present invention, control device 96 can dynamically change per user bandwidth allocations if a higher priority user or user device is sending/receiving network communications.

Assume, for example, User C is considered to have a higher priority than User A or User B. When User C connects to control device 96 with user device 94, User C can be allocated the full 500 Kbps for downloads specified in user profile 102. The download bandwidth allocations for User A and User B can be reduced to bring the total download allocation among all users to the 1000 Kbps capacity of user device 96. It should be noted that although the above examples discussed dynamic allocation of bandwidth in terms of downloads, upload bandwidth allocations can be similarly modified.

[0067] The program architecture of FIGURE 5 is provided by way of example only and dynamic bandwidth allocation on a per user basis can be implemented in any suitable programming manner known in the art. Additionally, control device 96 can dynamically change the bandwidth allocations to users according to any bandwidth shaping scheme known in the art. For example, control device 96 can adjust the bandwidth limits for users based on time of day, usage patterns or utilization averaging.

[0068] Dynamic bandwidth shaping based on time of day can allow control device 96 to provide particular users with additional bandwidth during peak hours. Allocating bandwidth based on usage patterns can allow a control device to set higher bandwidth limits for a user for a short period of time to account for, for example, bursty traffic. For short interactive traffic or short file transfers, this can give the user a high available

bandwidth. If the user continues to use a high amount of bandwidth, the user's bandwidth limits can be lowered so that user does not use too much of the control device's bandwidth capacity for too long. Utilization averaging allows for better fairness in allocation of excess bandwidth based upon measured utilization rates of competing users over a period of time. If the control device affords a users a higher bandwidth than specified in the user's user profile, the control device can dynamically lower the bandwidth allocation for that user to a level lower than specified in the user's user profile to allow other users additional bandwidth.

[0069] In summary, control device 96 can retrieve a set of user profiles (e.g., user profile 98, user profile 100 and user profile 102) for a set of users (User A, User B and User C) and establish upload and/or download bandwidth limits for each user based on the corresponding user profile for that user (e.g., user profile 98 for user A, user profile 100, for User B and user profile 102 for User C). The control device can regulate bandwidth usage for each user based on the bandwidth limit(s) established for that user. Additionally, the control device, according to one embodiment of the present invention, can update the bandwidth limit(s) for one or more of the users based on new users connecting to control device 96, excess capacity, utilization averaging, time of day or other factors known in the art.

[0070] FIGURES 6A-6D illustrate various embodiments of bandwidth shaping, using the example of User A and User C from FIGURE 5. In this case, it is assumed, for the sake of example, that User B is not connected to the control device. FIGURE 6A illustrates example bandwidth demands for User A and User C respectively. User A has a constant bandwidth demand (e.g., a file transfer), whereas User C has bursty traffic (e.g., web traffic). Both user's are initially allocated the same amount of bandwidth based on their respective user profiles (i.e., 500 Kbps). The dashed lines represents User A's bandwidth demand and User C's bandwidth demand for six time periods (period 122, period 124, period 126, period 128, period 130 and period 132). The control device, by way of example, but not limitation, has a maximum capacity of 1000 Kbps.

[0071] FIGURE 6B illustrates one embodiment of dynamically allocating bandwidth between User A and User C. The dashed lines represent the bandwidth demands from FIGURE 6A and the squared lines represent the bandwidth allocated to each user in a given time period. During all time periods User A is demanding more than the 500 Kbps specified in the User A's user profile. User C however, can be demanding more or less bandwidth in a given period. For example, in period 122, User C is demanding less bandwidth. Therefore, a control device can update the traffic control rules for User A and User C to allocate more bandwidth to User A and less bandwidth

to User C. During time period 126, both User A and User C are demanding at least 500 Kbps of bandwidth, making the total bandwidth demand greater than the available 1000 Kbps. The control device can again update the bandwidth allocations for User A and User C to limit User A and User C to the 500 Kbps specified in each user's profile. The control device can continue to allocate bandwidths to User A and User C as shown in FIGURE 6B. In one embodiment of the present invention, this can be done by updating the traffic control rules for each user, as described in conjunction with FIGURE 5.

[0072] FIGURE 6C illustrates another embodiment of bandwidth shaping using utilization averaging. As shown in FIGURE 6C, there will be periods (e.g., time period 128 and period 130) in which User C is able to use more than 500 Kbps, even though User A is continuing to have high demand. The allocation of bandwidth for each user can be based on a dynamic limit calculated for that user based on the user's utilization of bandwidth. The dynamic limit can be enforced whenever User A and User C compete for capacity (i.e., whenever their joint demands exceed 1000 Kbps). In one embodiment of the present invention, the dynamic limit for each user can be inserted into the traffic control rule for that user, whenever the user's compete for capacity. FIGURE 6D provides an example of the changing dynamic limit for User A, to produce the bandwidth allocation for User A shown in FIGURE 6C.

[0073] FIGURE 6D illustrates how a dynamic limit for User A can be changed in order to allow utilization averaging. In FIGURE 6D, the dashed line represents User A's demand, the solid line represents the dynamic limit for User A and the blocked line the represents the bandwidth allocated to User A. The demand and bandwidth allocation are the same as in Figure 6C for User A. The dynamic limit can be enforced for User A whenever User A and User C compete for bandwidth (i.e., when their demands exceed 1000 Kbps in this example). As shown in the example of FIGURE 6D, User A is permitted to take advantage of surplus bandwidth in periods 122 and 124 because User A and User C are not competing for capacity (see FIGURE 6C). The dynamic limit though not the bandwidth limit for User A is adjusted downward, based on the cumulative total of surplus bandwidth used.

[0074] At time period 126, the bandwidth allocation remains steady. During this time period, User C is only demanding 500 Kbps, so User A can be allocated the remaining 500 Kbps, even though the dynamic limit is below 500 Kbps. Both users can, therefore, be allocated 500 Kbps. When user C exceeds 500 Kbps (i.e., when User C competes for the excess bandwidth), the control device can enforce the dynamic limit for User A in time period 128. User A will see a throughput of less than 500 Kbps because the dynamic limit for User A is below 500 Kbps. During the time period in which User A is limited to less than 500 Kbps, the dynamic limit for User A can be slowly

increased. The pattern follows another decrease and increase in periods 130 and 132.

[0075] In one embodiment of the present invention, whether a dynamic limit for a user is adjusted up or down can be based on whether a user uses more or less bandwidth than the bandwidth limit specified in that user's user profile. In another embodiment of the present invention, whether a dynamic limit for a user is adjusted up or down can be based on whether a user uses more or less bandwidth than his or her proportional share of the available capacity (e.g., more or less than 50% if there are two users). The dynamic limit can be enforced for a user by modifying the traffic control rule for that user to include the dynamic limit as a bandwidth limit for the user. The traffic control rule for each user can be updated on an arbitrary schedule to enforce the changing dynamic limits.

[0076] It should be noted that, according to one embodiment of the present invention, a user who sustains a long period of high demand should not be subject to an ever-decreasing allocation of bandwidth. This could result in severe attenuation of the user's bandwidth the moment other users have an increased demand. In one embodiment of the present invention, two approaches can be used separately or in combination to remedy this: a lower limit on the dynamic bandwidth, and a rolling window for cumulative measurement. A lower limit would establish that a user's bandwidth allocation never drop below some

point, for instance half of the rate specified in the user's user profile. A rolling window defines a limited period of time, say ten minutes, for the cumulative measurement. This would mean that the user's bandwidth allocation calculated as though prior to the last ten minutes, no surplus bandwidth had been used, and the bandwidth allocation for the user had been set as specified in the user's user profile.

[0077] It should be noted that above examples of dynamic bandwidth allocation are provided by way of example only. A control device can dynamically calculate bandwidth allocations for users according to any bandwidth shaping scheme known in the art. Moreover, updated bandwidth limits can be provided to a traffic conditioning module in any suitable manner, as would be understood by those of ordinary skill in the art.

[0078] FIGURE 7 is a diagrammatic representation of one embodiment of a control device 141 that can provide user based provisioning of bandwidth. For the purposes of example, control device 141 can comprise a main bus 142, a main processor 144, a primary storage medium 146, a secondary storage controller 148, a storage medium 150, a user side network interface 152 and a controlled network network interface 154. The network interfaces can include Ethernet interfaces, fibre channel interfaces, T1 interfaces, wireless interfaces or other network interfaces known in the art. Other devices may be connected to or be part of such a control device include,

by way of example, but not limitation, controllers, a display, a mouse, a keyboard, and so forth. Additionally, control device 140 can include additional interfaces to communicate to additional networks using various protocols and can include interfaces for administrative functions.

[0079] The main processor 144 communicates with the other components by way of the main bus 142. This main processor 144 can be a general purpose processor, a limited processor such as an ASIC or microcontroller, or any other instruction execution machine. The primary storage 146 can provide transient memory or storage space for use by programs executing on the main processor 144. The main processor 144 communicates with the primary storage in any manner known in the art.

[0080] The secondary storage controller 148 connects a storage medium 150 such as a hard drive, CD-ROM, floppy, tape drive, optical storage medium, memory or other storage device to the main processor 144 by way of the main bus 142. The main processor 144 communicates with the secondary storage controller 148 by way of the main bus 142, and the secondary storage controller 148 is used to read and /or write the storage medium 150 on behalf of the main processor 144.

[0081] Control device 141 may communicate with other computing devices (e.g., user devices, network servers, etc.) by way of networks using network interfaces (e.g., user side

network interface 152 and controlled network network interface 154 or other network interface). Computer instructions running on the main processor may then access other computers across the network in any of the conventional ways, e.g. by executing "protocols" which affect the transmission and reception of protocol data units, packages, etc. over the data transmission network.

[0082] In one embodiment of the present invention, storage medium 150 can store a set of computer instructions 156 that are executable by processor 144. During execution, portions of computer instructions 156 and data can be stored in primary storage 146, as would be understood by those of ordinary skill in the art. Processor 144 can execute computer instructions 156 to allocate bandwidths on a per user basis and enforce the bandwidth allocations.

[0083] Although shown as a standalone device in FIGURE 7, control device 141 may be integrated with and share components with other devices such as routers, servers, hubs or other devices known in the art. Additionally, computer instructions 156 can be distributed across multiple storage media and can be executed by multiple processors. One example of an exemplary control device is the Rocksteady NSA Server, from Rocksteady Networks, Inc. of Austin, Texas.

[0084] Control device 141, as described in the Access Control Application, can also provide any arbitrary services

known in the art, including, but not limited to, web server functions, DHCP client for negotiation with ISPs, DHCP server to assign IP addresses to user devices, kernel based packet filtering and stateful inspection, IP sharing, NATplus, port redirection, information and attack logging, automatic updating, VPN masquerade, remote support and configuration, name server configuration and/or web content filtering. User profiles can be used by the control device to govern provisioning of network access on a user specific basis. By way of example, but not limitation, a user profile can contain attributes to specify upload and download bandwidth allocations for a user, firewall settings, whether the user can use transient VPNs, whether the user can use streaming services or voice over IP services, whether the user should be permitted to perform video teleconferencing, whether the control device should perform virus scanning or worm detection for the user, whether the user can utilize print services, surcharges for services or other settings.

[0085] While the present invention has been described with reference to particular embodiments, it should be understood that the embodiments are illustrative and that the scope of the invention is not limited to these embodiments. Many variations, modifications, additions and improvements to the embodiments described above are possible. It is contemplated that these variations, modifications, additions and improvements fall within the

scope of the invention as detailed in the following claims.